BMC Genomics

RESEARCH

Open Access

Prediction of lipase-specific foldase-dependence in bacterial lipase subfamilies I.1 and I.2

Takahiro Hioki[1*]

Abstract

Background Most bacterial lipases in subfamily I.1/I.2 depend on a specific chaperone protein, lipase-specific foldase (Lif), for folding into their active form. In contrast, several Lif-independent lipases have been reported in subfamily I.1. Lif-independent lipases have the potential to be industrially useful owing to their ease of heterologous expression; however, no method has been reported to predict Lif-dependence for an arbitrary lipase. In this study, we comprehensively estimated the Lif-dependence of subfamily I.1/I.2.

Results To estimate Lif-dependence, we comprehensively analyzed the presence or absence of Lif genes in the genomes of bacteria from which the lipases were derived and integrated the results with those of phylogenetic analysis. We identified a range of lipases from the *Pseudomonas fragi/Proteus vulgaris* clade, which contained all known Lif-independent lipases and were enriched for lipases that did not coexist with Lif. Sequences and structural features conserved in the *P. fragi/P. vulgaris* clade and other lipases were identified, and the residues involved in Lif-dependence were inferred. Furthermore, we identified the *Pseudoalteromonas shioyasakiensis* clade, which is phylogenetically distinct from the *P. fragi/P. vulgaris* clade, as having no Lif in the genome of the bacterium from which the lipase was derived. The *P. shioyasakiensis* clade lipase, PliLip, was heterologously expressed in *Escherichia coli* in an active form, independent of Lif.

Conclusions In this study, we developed a method to predict Lif-dependence in any lipase belonging to subfamily I.1/I.2 and comprehensively extracted putative Lif-independent lipases from public databases. This study contributes to expand the diversity of industrially available Lif-independent lipases and provides fundamental insights into the evolution of lipases and Lif.

Keywords Bacterial lipase, Lipase-specific foldase, Molecular chaperone, Phylogenetic analysis, Heterologous expression

Background

Lipases are enzymes that hydrolyze ester bonds in water-insoluble carboxylic esters and are produced by various species. Microbial lipolytic enzymes are involved in diverse biological functions including lipid metabolism and pathogenicity [1]. They are also important in biotechnology, where they are used industrially for a wide range of applications, including detergents, fat, oleochemicals, and biodiesel production as well as for enantioselective synthesis [2].

Arpigny et al. classified bacterial lipolytic enzymes into eight families based on comparisons of amino acid sequences and biological properties, with the largest family, Family I, divided into six subfamilies [3]. Since then,

*Correspondence:
Takahiro Hioki
hioki.takahiro@kao.com
[1] Biological & Material Science Research, Kao Corporation, 1334 Minato, Wakayama, Wakayama 640-8580, Japan

many classifications have been added for newly discovered enzymes, and 35 families and 11 subfamilies of Family I have been recently proposed [4]. Family I includes a group of lipases called "True Lipases," which are highly specific for long-chain triglycerides that are insoluble in water. Subfamilies I.1 and I.2 of Family I include *Pseudomonas* and *Burkholderia* lipases, respectively, which are the most important bacterial lipases in the industry. Subfamilies I.1 and I.2 show higher sequence homology with each other than with other subfamilies and are evolutionarily related [3, 5].

Many lipases in subfamilies I.1 and I.2 require a specific chaperone protein called a lipase-specific foldase (Lif) for folding into an enzymatically active conformation [6]. These lipases are considered to be transferred to the periplasm via the Sec pathway, activated by interaction with plasma membrane-anchored Lif, and secreted across the outer membrane into the extracellular space by a type II secretion mechanism [7]. Lif is speculated to act as a chaperone to help overcome the energy barriers in the productive folding pathway [8]. Lif specifically activates lipases of the same or closely related species. *Pseudomonas aeruginosa* Lif cannot substitute for *Burkholderia glumae* or *Acinetobacter calcoaceticus* Lif but has been shown to activate a lipase from the closely related species, *Pseudomonas alcaligenes* [9, 10]. The lipase gene and its cognate Lif gene often form an operon [6].

Many lipases, which are classified into subfamilies I.1 and I.2, are derived from human and plant pathogenic bacteria. Especially for applications involving direct exposure of humans to enzymes, such as in detergents, production using heterologous expression systems in hosts with widely recognized safety is preferable. Using a heterologous host with a high protein production capacity is also beneficial from a cost perspective. However, heterologous expression of lipases from subfamilies I.1 and I.2 is difficult because of their Lif-dependence. The expression of these lipases in *Escherichia coli* yields inclusion bodies of inactive lipases [11–14]. Activation of these inactive enzymes requires in vitro refolding in the presence of Lif, which is unsuitable for industrial production. Co-expression of lipase and Lif yields active lipases *in vivo* [15–17]. However, in the expression of *P. aeruginosa* LipA in *E. coli*, most lipases are reported to be expressed in the insoluble fraction, even with Lif co-expression [17]. Thus, inexpensive production of subfamily I.1 and I.2 lipases by heterologous expression remains challenging.

Some enzymes classified as subfamily I.1, such as lipases from *Pseudomonas fragi* and *Proteus vulgaris*, have been reported to fold into their active form during *E. coli* heterologous expression in a Lif-independent manner [5, 18–21]. All known Lif-independent lipases from subfamily I.1 are evolutionarily related and form an independent clade (*P. fragi*/*P. vulgaris* clade) [5, 22]. No Lif genes have been found in the vicinity of lipase genes in the genome of bacteria from which these lipases are derived. In fact, the soluble expression of *P. fragi* lipase was not improved by co-expression with *P. aeruginosa* Lif [21]. The known lipases of the *P. fragi*/*P. vulgaris* clade have two unique features in their amino acid sequences compared with those of other subfamily I.1 lipases. First, the lipases of subfamilies I.1 and I.2 usually possess an N-terminal secretion signal, whereas none of the known lipases belonging to the *P. fragi*/*P. vulgaris* clade possess a distinct secretory signal. Nevertheless, these lipases are released extracellularly by their natural producers, suggesting extracellular transport via lysis or an unknown secretory mechanism [18, 19]. Second, known lipases belonging to the *P. fragi*/*P. vulgaris* clade lack the intramolecular disulfide bonds that are widely conserved in subfamilies I.1 and I.2 [5, 21].

Lif-independent lipases have the potential to be industrially useful because they can be heterologously expressed in their active form without the need for Lif co-expression. LipC12 identified in the metagenome exhibits properties suitable for biotechnological applications such as excellent stability under high-temperature conditions and in organic solvents, high activity at room temperature, high heterologous expression, and solubility [5, 23]. Although the difficulty in heterologous expression has hindered protein engineering studies on subfamily I.1 and I.2 lipases, efficient protein engineering is possible with Lif-independent lipases. For instance, 13 mutations were introduced into the lipase from *Proteus mirabilis*, which increased its methanol tolerance for biodiesel production [24].

The development of diverse Lif-independent lipases is increasingly important for biotechnology; however, no method is currently available to predict the Lif-dependence of any lipase. Previous studies have examined the sequences and structural characteristics of lipases classified in the *P. fragi*/*P. vulgaris* clade [5, 25], only a small number of experimentally validated sequences have been used and the general characteristics of this clade have not been fully clarified. Further, in subfamilies I.1 and I.2, the only known clade of Lif-independent lipases is the *P. fragi*/*P. vulgaris* clade (classified in subfamily I.1).

In the genomes of bacteria with genes encoding Lif-dependent lipase, Lif genes should be present in the vicinity of the lipase genes in most cases or at least somewhere on the genome. Therefore, we considered that Lif-dependence can be estimated by the presence or absence of Lif genes on the genome and thus, comprehensively analyzed Lif coexistence for lipases in public databases.

Table 1 Previous studies on the Lif-dependence of lipase

Subfamily	Lif-dependence	Accession	Organism	Reference
I.1	Yes	D50587	*Pseudomonas aeruginosa*	[3, 34]
I.1	Yes	X16945	*Vibrio cholerae*	[3, 35]
I.1	Yes	X80800	*Acinetobacter calcoaceticus*	[3, 36]
I.1	Yes	ABE74226	*Psychrobacter cryohalolentis* K5	[37]
I.1	Yes	AAG47649	*Pseudomonas* sp. SW-3	[38]
I.1	No	AF031226	*Pseudomonas fluorescens* C9	[3, 39]
I.1	No	X14033	*Pseudomonas fragi*	[3, 18]
I.1	No	U33845	*Proteus vulgaris* K80	[3, 20]
I.1	No	AIZ05356	*Yersinia enterocolitica* KM1	[40]
I.2	Yes	X70354	*Burkholderia glumae*	[3, 8, 41]
I.2	Yes	AAR13272	*Ralstonia sp.* M1	[42]

Methods

Phylogenetic analysis

The known subfamily I.1/I.2 lipases [3] from *Pseudomonas aeruginosa* (GenBank accession No. D50587), *Pseudomonas fluorescens* C9 (GenBank accession No. AF031226), *Vibrio cholerae* (GenBank accession No. X16945), *Acinetobacter calcoaceticus* (GenBank accession No. X80800), *Pseudomonas fragi* (GenBank accession No. X14033), *Pseudomonas wisconsinensis* (GenBank accession No. U88907), *Proteus vulgaris* K80 (GenBank accession No. U33845), *Burkholderia glumae* (GenBank accession No. X70354), and *Burkholderia cepacia* (GenBank accession No. M58494) were used as queries; a Protein BLAST search [26] against the NCBI nr protein database [27] yielded 10218 lipase sequences, excluding duplicates, with an e-value = 1e-20. From these sequences, 7332 amino acid sequences annotated to the RefSeq genome (accession numbers starting with WP_), with a sequence length of 250–400 residues, and a query cover >60% were extracted. In the genome from which each lipase was derived, the availability of a 5 kb genomic sequence both upstream and downstream of the lipase ORF was examined using Entrez [28] and 6543 lipase sequences for which genomic sequences were available were extracted. cd-hit [29] was used to cluster sequences with 70% sequence identity and 329 lipase sequences were obtained. Multiple sequence alignment was performed using MAFFT-DASH [30] (Method: FFT-INS-i) with 11 known lipase sequences (Table 1), along with yeast-derived sterol esterase (UniProt accession No. P34163) and gorilla-derived lipase (UniProt accession No. G3QN03) used as outgroups. Positions with more than 10% of sequences being gaps, and sequences with more than 20 gaps in the remaining multiple sequence alignment were deleted. Finally, 327 sequences, comprising 314 lipase sequences, 11 known lipase sequences, and two outgroup sequences, were subjected to phylogenetic analysis. Phylogenetic analysis was performed using IQ-TREE 2 [31]. The phylogenetic tree was visualized using iTOL [32]. Secretory signals were analyzed using SignalP 6.0 [33].

Lif coexistence analysis

Of the sequences subjected to phylogenetic analysis, 314 were used for Lif coexistence analysis, excluding known lipases and outgroups. Using Entrez, all ORFs within 5 kb upstream and downstream of the lipase ORFs were extracted from the bacterial genome from which each lipase was derived, yielding a total of 2642 amino acid sequences. The raw hidden Markov model (HMM) of the proteobacterial lipase chaperone protein (PF03280) was obtained from the Pfam [43] database, and a homology search was performed for 2642 amino acid sequences with PF03280 using HMMER [44]. In total 250 Lif sequences that scored above the Pfam classification threshold of 25.7 were extracted. These sequences were then subjected to multiple sequence alignments using MAFFT-DASH (Method: FFT-INS-i) to generate HMM profiles (LifHMM). A homology search with the Lif-HMM was performed on the 2642 amino acid sequences using HMMER, and 252 sequences that scored > 25.7 were determined as Lif (Table S1).

For 68 lipase genes that did not coexist with Lif within 5 kb of the genome, the ID of the genome assembly, from which each lipase was derived, was obtained using Entrez, and all ORFs in each assembly were extracted using the ncbi-genome-download tool [45] for a total of 246953 amino acid sequences. A homology search with the LifHMM was performed, and sequences scoring >25.7 were determined as Lif.

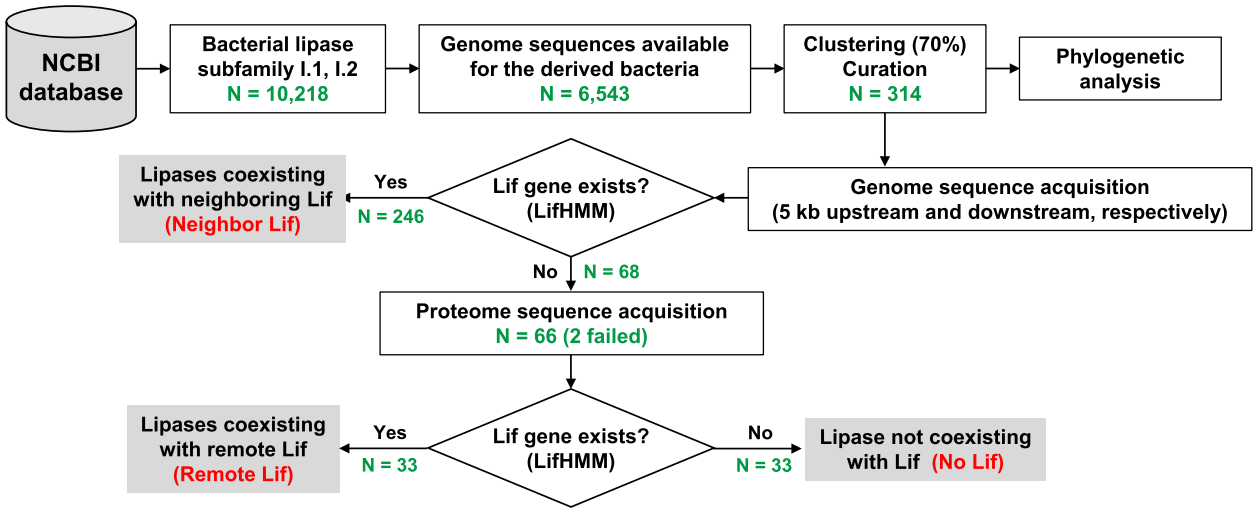

Fig. 1 Overview of Lif coexistence analysis. (i) A homology search using known subfamily I.1/I.2 lipases as queries was performed to extract subfamily I.1/I.2 lipases in the NCBI nr database. (ii) Only lipase sequences for which the genome sequence of the derived bacteria was available were extracted and clustered with 70% identity. (iii) LifHMM was defined and examined for the presence of Lif genes in the vicinity of the lipase genes. (iv) For lipase genes with no Lif genes detected in the vicinity, Lif genes were searched in the entire proteome of the derived bacteria. (v) Phylogenetic analysis was performed on the same set of lipase sequences. Green letters indicate the number of sequences at each step

Analysis of the *P. shioyasakiensis* clade

In total, 85 lipase sequences were obtained by recovering all sequences in the cluster with 70% sequence identity for each of WP_062566573.1, and its surrounding sequences in the phylogenetic tree (WP_149982606.1, WP_138549921.1, WP_193332450.1, WP_015430253.1, and WP_209113137.1). The 28 lipase sequences obtained by clustering again with 95% sequence identity using cd-hit were subjected to phylogenetic and Lif coexistence analyses using the same method described above.

Sequence comparison

For sequences classified as subfamily I.1 in the phylogenetic analysis, the sequences before clustering were extracted and re-clustered at 90% identity. Secretory signal sequences were removed using SignalP 6.0. Multiple sequence alignment was performed using the MAFFT-DASH (Method: G-INS-i) and the alignment was corrected visually. The alignments were then divided into two groups: the *P. fragi/P. vulgaris* clade and others (called Other clades). Positions with gaps in more than 70% of the sequences in both groups were trimmed. For each group, a sequence logo was created using WebLogo [46]. Only positions in which the same amino acid was conserved in more than 80% of the sequences in either group and the conserved amino acid was present in only 10% or less of the sequences in the other group were extracted, and sequence logos were created. PAL (UniProt accession No. P26876, excluding signal sequences) and PML (UniProt accession No. B4EVM3) were used to number the sequence logos. Sequence logos were also created for sequences classified as subfamily I.2 in the phylogenetic analysis.

Protein complex structure prediction

ColabFold v1.5.2 was used for protein structure prediction [47, 48]. Sequences with the secretory signal sequence removed were used for the prediction. To predict the structure of the complex of PAL and the cognate Lif, the sequence after residue 59 of Lif (UniProt accession No. Q01725), excluding the transmembrane and linker regions, was used and prediction was performed with relax = 1 and recycle = 3. Complex structure prediction of the hit sequences of LifHMM and neighboring lipases was performed with no relax, recycle = 1. The Non-Bond Interaction method in Discovery Studio 2022 (Dassault Systèmes, Vélizy-Villacoublay, France) was used to analyze residue-residue interactions in the protein structure.

Results and discussion

Lif coexistence analysis

The flow used for the comprehensive analysis of Lif coexistence is shown in Fig. 1. Lipases classified as subfamilies I.1 or I.2 by Arpigny et al. were used as query sequences [3] and homologous sequences were collected from the NCBI database. Only lipase sequences with available bacterial genome sequences were selected for clustering and curation, and finally, 314 lipase sequences were analyzed for coexistence with Lif. Lif located within 5 kb upstream

or downstream of the lipase gene in the genome was defined as Neighbor Lif, and that located further away was defined as Remote Lif.

First, Neighbor Lif was analyzed. All open reading frames (ORFs) within 5 kb upstream and downstream of each lipase gene were extracted, yielding a total of 2642 amino acid sequences. HMMER [44] was used to search for Lif genes. We then tested whether the Lif gene could be identified by analyzing whether it was classified into the Proteobacterial lipase chaperone protein (PF03280) family of the Pfam database [43]. A homology search of 2642 sequences yielded 252 sequences homologous to PF03280, of which 250 sequences exceeded the classification threshold (a score of 25.7). However, two sequences with scores below the threshold (WP_089605658.1: score 22.2, WP_140626750.1: score 25) were both located next to the lipase gene in the genome and were predicted to form a complex with a neighboring lipase showing high confidence in the structure prediction (Table S2). Therefore, these are highly likely to act as Lif. Overall, the 250 sequences classified as PF03280 in the above analysis were highly likely to function as Lif because they are located near the lipase gene and show high homology with known Lifs. Therefore, to search for Lif genes with higher sensitivity, we generated a new hidden Markov model (HMM) profile (LifHMM) from 250 sequences and again performed a homology search. This yielded 258 sequences, of which 252 sequences had a score of 25.7 or higher. These were consistent with the 252 sequences identified in the search using PF03280 HMM, with the two sequences below the threshold in PF03280 also showing high scores (WP_089605658.1: score 51.2; WP_140626750.1: score 134.8). The sequence with the highest score among the six sequences that hit below the threshold, WP_219123567.1 (score: 15.3), was not predicted to form a specific complex with a neighboring lipase and was not likely to act as a Lif (Table S2). Finally, 252 sequences were determined to be Lif genes. These were mapped to lipase sequences, and 246 of 314 lipase sequences were found to contain Neighbor Lif genes (the corresponding lipase sequences were labeled "Neighbor Lif"). For 68 lipases that did not have a Neighbor Lif gene, Lif was searched throughout the proteome of each source bacterium. The LifHMM search revealed Lif in 33 of the 66 derived bacteria for which proteome sequences were successfully obtained (the corresponding lipase sequences were labeled "Remote Lif"). No Lif was detected in the remaining 33 proteomes (the corresponding lipase sequences were labeled "No Lif").

Phylogenetic analysis and signal prediction were performed by adding lipases with known Lif-dependence (Table 1) as well as outgroups to the analyzed sequences. The results of Lif coexistence and signal analysis were then integrated into a phylogenetic tree (Fig. 2, Fig. S1, Table S3).

Overview of the analysis results

Two previously reported lipases classified in subfamily I.1 and nine previously reported lipases classified in subfamily I.2 were dichotomized into different clades with high confidence (SH-aLRT/UFboot = 99.9/100). Therefore, all lipases used for the analysis could be clearly classified into subfamily I.1 or I.2.

Most known Lif-dependent lipases cluster with Lif in the genome. One exception is the *lipC* gene of *P. aeruginosa*, which has no Lif genes in its vicinity and depends on *lipB*, a neighboring Lif gene of another lipase, *lipA*, for its active expression [49]. Many lipases in this analysis were also found to coexist with Neighbor Lif genes, whereas lipases coexisting only with Remote Lif genes were scattered throughout the phylogenetic tree.

All lipases from subfamily I.2 coexisted with Lif, consistent with previous reports. However, some subfamily I.2 lipases did not have typical secretory signals, and some only coexisted with Remote Lif, indicating diversity in the localization and expression patterns of subfamily I.2 lipases.

A group of lipases with no Lif detected in the genome (No Lif) formed an independent clade (*P. fragi*/*P. vulgaris* clade) with all known Lif-independent lipases, except for one sequence described below. This supports the validity of the analysis. The fact that all lipases coexisting with Remote Lif in this clade also coexisted with other lipase-Lif clusters (Table S4) suggests that they may be as Lif-independent as the other lipases of this clade. Experimental verification is needed to clarify the Lif-dependence of these lipases, such as examining the presence or absence of lipase activity under heterologous expression of lipases alone. All lipases of the *P. fragi*/*P. vulgaris* clade are predicted to have no secretory signal, consistent with the findings of previous reports [18, 19].

Sequence comparison of the *P. fragi*/*P. vulgaris* clade and other clades

To characterize the sequence features of the *P. fragi*/*P. vulgaris* clade and other subfamily I.1 lipases, we dichotomized subfamily I.1 into the *P. fragi*/*P. vulgaris* clade and other clades (hereafter, Other clades) and compared their sequence conservation (Fig. S2). The residue positions where some amino acids were highly conserved in only one of the groups were extracted and are shown in Fig. 3A. All residue numbers indicated below correspond to the residue numbers of the lipase from *P. aeruginosa* PAO1 strain (PAL).

C183–C235, which are highly conserved in the Other clades, contribute to the structural stabilization of the

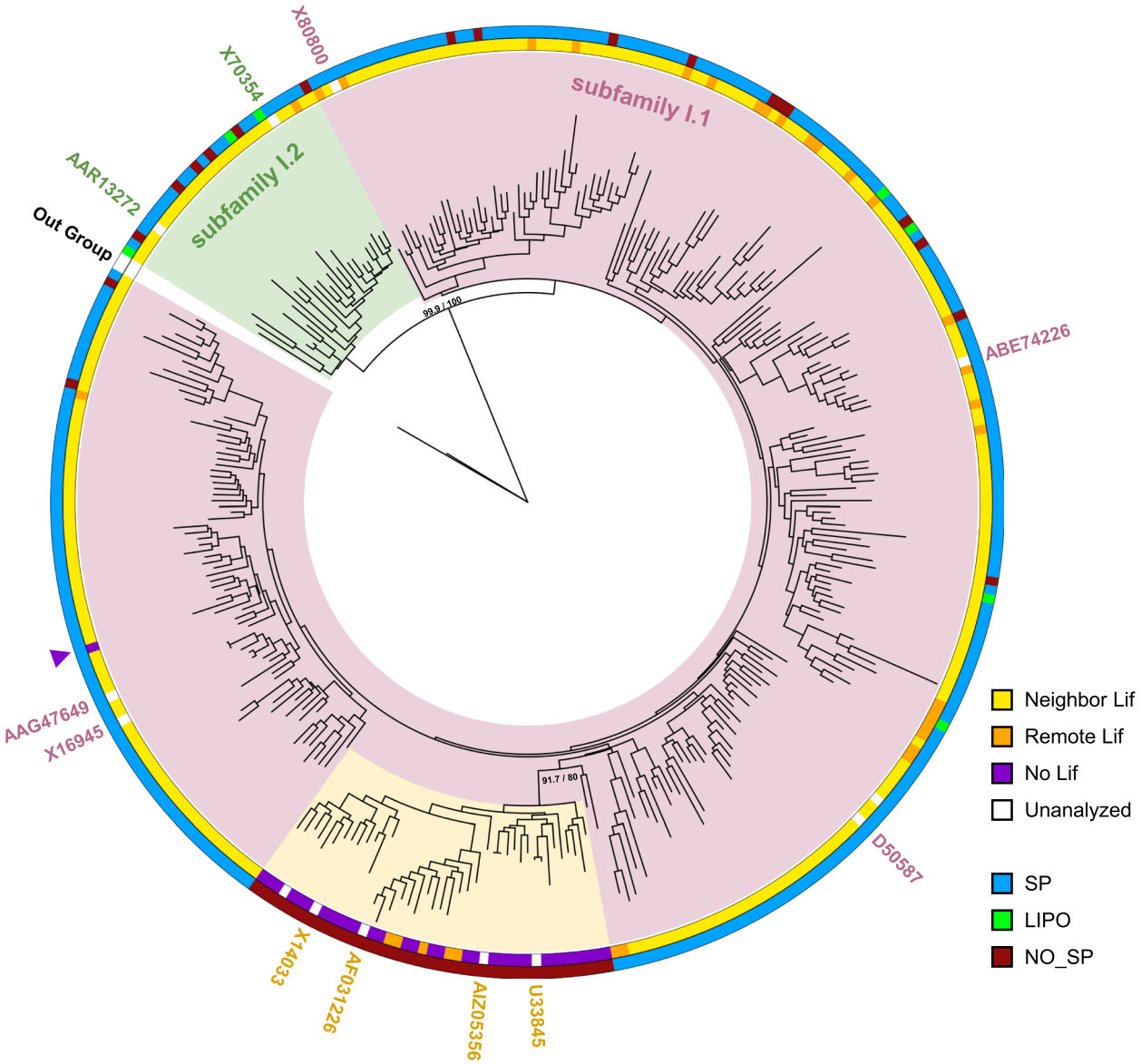

Fig. 2 Phylogenetic tree and Lif coexistence of subfamily I.1 and I.2 lipases. In the phylogenetic tree, the background is pink for subfamily I.1, green for subfamily I.2, and yellow for the *P. fragi*/*P. vulgaris* clade. Of the two colored labels outside the phylogenetic tree, the inner label represents the result of the Lif coexistence analysis: yellow for Neighbor Lif, orange for Remote Lif, purple for No Lif, and white for unanalyzed sequences. The outer labels represent the results of signal prediction: light blue for SP (Sec/SPI), yellow-green for LIPO (Sec/SPII), and burgundy for sequences with no signal (OTHER). Outside the colored labels are the accession numbers of the existing sequences. WP_062566573.1, classified under the *P. shioyasakiensis* clade, is indicated by a purple triangle

active enzyme by forming disulfide bonds [50]. This cysteine residue was not conserved in the *P. fragi*/*P. vulgaris* clade; instead, a conserved E225 residue was presumed to contribute to structural stabilization [25]. In the crystal structure of *Proteus mirabilis* lipase (PML), E225 is suggested to stabilize helix η1 by forming a hydrogen bond with the main chain of the non-conserved residue S235, contributing to the correct orientation of the active residue D229. Consistent with previous reports, this analysis indicated that C183–C235 is not conserved in the *P. fragi*/*P. vulgaris* clade, and that E225 is highly conserved. Furthermore, E113, which has not been previously described, is highly conserved only in the *P. fragi*/*P. vulgaris* clade. The side chain of E113 forms a hydrogen bond with the main chain of E225 in the PML crystal structure, suggesting the importance of the E113-E225-X235 hydrogen bond network for stabilizing the active center D229 in the correct orientation (Fig. 3B).

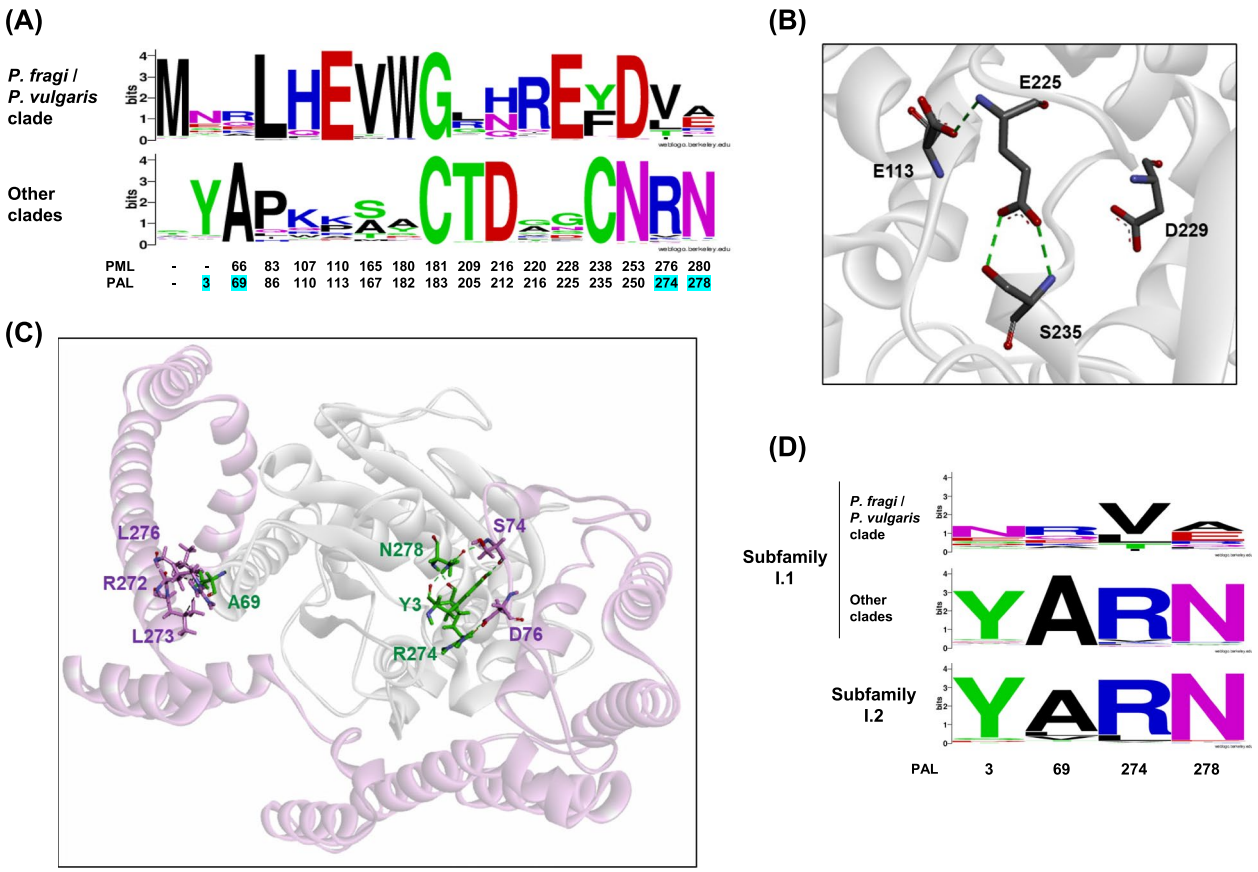

Fig. 3 Comparison of the *P. fragi*/*P. vulgaris* clade with other clades. **A** Sequence comparison of the *P. fragi*/*P. vulgaris* clade with other clades of subfamily I.1. Residue numbers of the corresponding amino acid residues in PML and PAL are shown below the logo. In the predicted complex structure of PAL and its cognate Lif, the residue numbers predicted to interact with Lif are shown in a light blue background. **B** The E113-E225-S235 hydrogen bond network in PML. The PML crystal structure (PDB ID: 4 gw3) is represented by a ribbon model, with E113, E225, S235, and the active center D229 represented by sticks and hydrogen bonds represented by green dashed lines. **C** Complex structures of PAL and its cognate Lif predicted using AlphaFold2. PAL is represented by a white ribbon model and Lif by a purple ribbon model. Y3, A69, R274, and N278 of PAL are represented by green sticks, and residues on Lif predicted to interact with them are represented by purple sticks. **D** Sequence comparison of the *P. fragi*/*P. vulgaris* clade, other clades of subfamily I.1, and subfamily I.2. Sequences of residue positions corresponding to Y3, A69, R274, and N278, predicted to interact with Lif in PAL, were compared in the three groups. Residue numbers of the corresponding amino acid residues in PAL are shown below the logo

The coordination of calcium ions plays an important role in maintaining catalytic histidine (H251 in PAL) at the correct position in the active site [25, 51]. The side chains of T205, D212, and N250, which are highly conserved only in Other clades, interact with water molecules that coordinate calcium ions in the PAL crystal structure [51], suggesting that T205/D212/N250 is a common strategy for stabilizing the coordination of calcium ions in Other clades. Further, D250, which interacts with water molecules that coordinate calcium ions in the PML crystal structure, is specifically and highly conserved in the *P. fragi*/*P. vulgaris* clade, consistent with previous reports [25]. The presence of a characteristic loop (residues 201–208 of PML), longer than that in PAL, as previously mentioned [25], was also conserved, but its sequence and length were less conserved.

The crystal structures of lipases from *B. glumae* and *Acinetobacter baumannii* in complex with their cognate Lif have been reported, and the residues at which the lipase and Lif interact with each other have been clarified [52, 53]. Here, we analyzed the residues that interact with Lif, which are selectively conserved only in Lif-dependent lipases, to identify residues that are particularly important in the interaction with Lif. Y3, A69, R274, and N278, which are highly conserved only in Other clades, are predicted by AlphaFold2 to interact with Lif in the complex structure of PAL and cognate Lif [47] (Fig. 3C). Y3-R274-N278 were predicted to form a cluster, with the side chains of each residue forming hydrogen bonds with

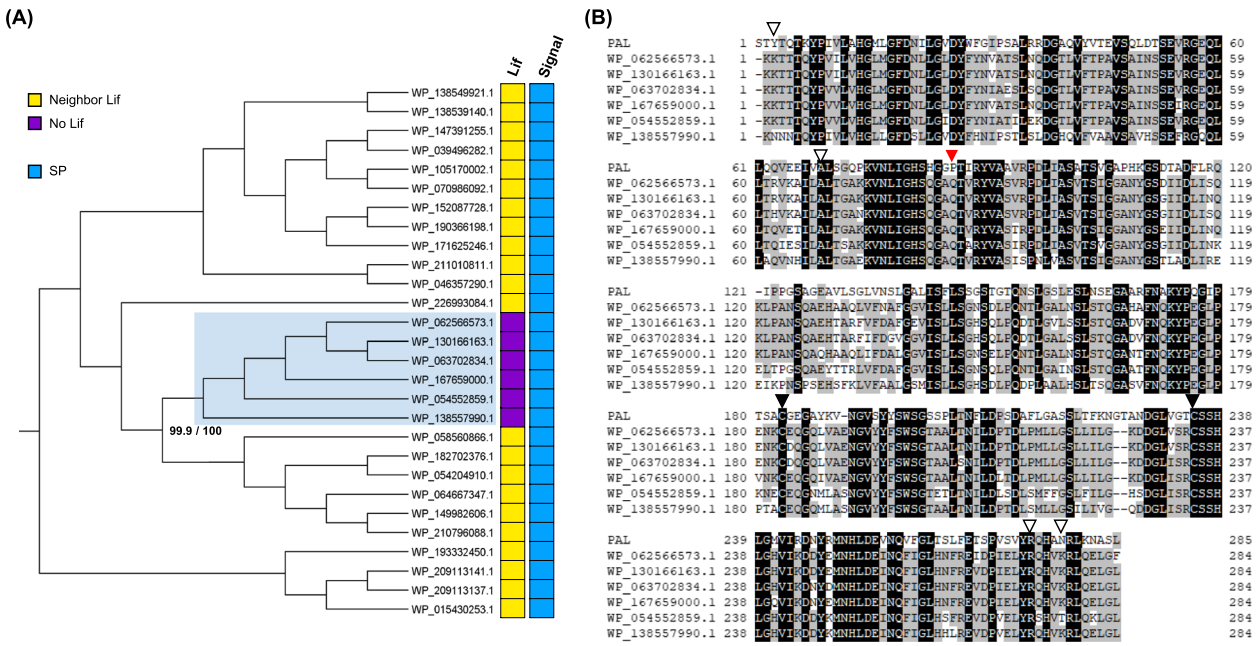

Fig. 4 Identification of the *P. shioyasakiensis* clade. **A** Phylogenetic and Lif coexistence analysis of sequences around WP_062566573.1. Of the two color labels on the right side of the phylogenetic tree, the left label represents the result of Lif coexistence analysis, yellow indicates Neighbor Lif and purple indicates No Lif. The right label represents the result of signal prediction, and light blue indicates SP(Sec/SPI). *P. shioyasakiensis* clade was defined based on the result of Lif coexistence analysis, and the background of the phylogenetic tree was colored light blue. **B** Sequence alignment of PAL with six sequences belonging to *P. shioyasakiensis* clade. Y3, A69, R274, and N278, which are predicted to interact with Lif in PAL, are indicated by white triangles, and C183 and C235, which form intramolecular disulfide bonds, by black triangles, P86, which corresponds to a residue suggested to be involved in Lif-dependency in Lif-dependent lipase from *Pseudomonas* sp. KFCC10818, is indicated by a red triangle. Residues with the same amino acid conserved in all sequences are colored in black, and residues with the same amino acid conserved in more than half of the sequences are colored in gray on the background

S74, D76, and S74 of Lif, respectively. The methyl group in the side chain of A69 was predicted to interact hydrophobically with the R272, L273, and L276 of Lif. Four residues, Y3, A69, R274, and N278, are highly conserved in subfamily I.2 but are specifically not conserved only in the *P. fragi/P. vulgaris* clade, which did not require interaction with Lif (Fig. 3D). These results suggest that these residues play a particularly important role of in the interaction with Lif.

Pseudoalteromonas shioyasakiensis clade

Other than the *P. fragi*/*P. vulgaris* clade, only lipase WP_062566573.1, from *Pseudoalteromonas shioyasakiensis*, had no Lif detected in the genome (Fig. 2, purple triangle). The same analysis was performed for the surrounding sequences by changing the clustering from 70% to 95%. The results showed that Lif was not detected in any of the six sequences, including WP_062566573.1, which formed a small clade (defined as the *P. shioyasakiensis* clade) (Fig. 4A). This suggests that the *P. shioyasakiensis* clade lipases are a novel group of Lif-independent lipases.

The alignment of the *P. shioyasakiensis* clade with the PAL clade is shown in Fig. 4B. Two cysteine residues that form disulfide bonds in PAL are also conserved in the *P. shioyasakiensis* clade. Among Y3, A69, R274, and N278, which were suggested to be important for the interaction with Lif, Y3 and N278 were not conserved in the *P. shioyasakiensis* clade. In a Lif-dependent lipase from *Pseudomonas* sp. KFCC10818 belonging to subfamily I.1, the substitution of proline with glutamine in a residue corresponding to P86 of PAL has been reported to allow heterologous expression in an active form independent of Lif [54]. This position of the residue is highly conserved, with proline in the Other clades and leucine in the *P. fragi*/*P. vulgaris* clade, whereas glutamine is conserved in the *P. shioyasakiensis* clade. This suggests that glutamine at position 86 in *P. shioyasakiensis* clade may contribute to Lif-independence. Notably, the *P. shioyasakiensis* clade possessed a secretory signal, unlike the *P. fragi*/*P. vulgaris* clade.

Finally, the heterologous expression of the lipase PliLip (WP_054552859.1), belonging to the *P. shioyasakiensis* clade, and for which no Lif was detected in the genome (Fig. 4A), was examined in *E. coli*. The strain transformed

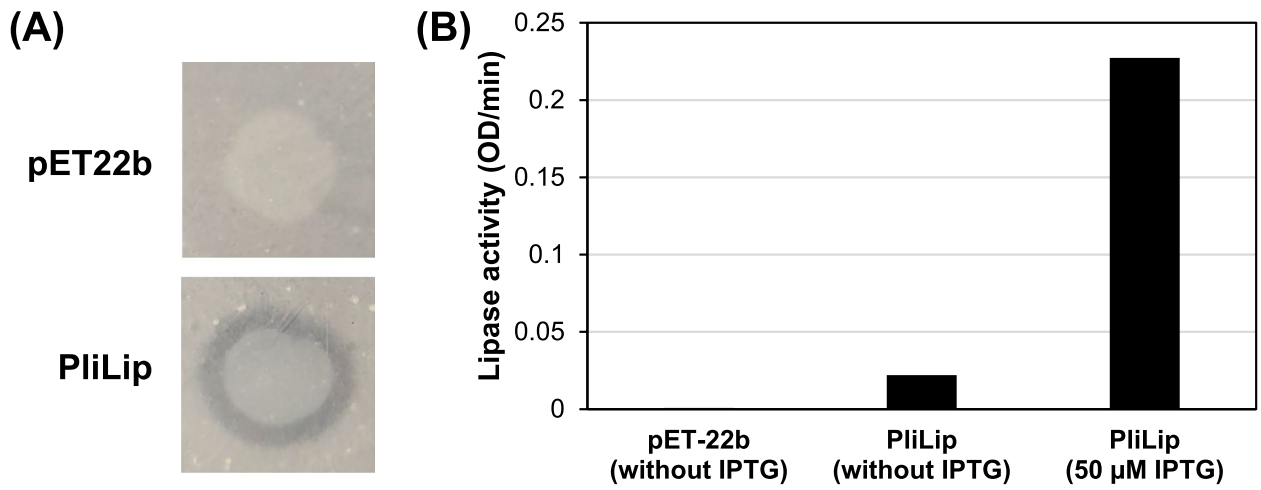

Fig. 5 Heterologous expression of PliLip in *E. coli*. **A** Hydrolysis halo formed by PliLip-producing cells in the tributyrin plate assay. **B** Measurement of lipase activity of culture supernatant with p-nitrophenyl butyrate as a substrate

with the vector encoding PliLip formed halos on tributyrin agar medium (Fig. 5A). IPTG induction also induced pNP-butyrate degradation activity in liquid culture (Fig. 5B). These results confirmed that PliLip is a Lif-independent lipase and identified the *P. shioyasakiensis* clade as a novel group of Lif-independent lipases.

Conclusions

The Lif-dependence of a wide range of lipases classified into subfamilies I.1 and I.2 was inferred by analyzing Lif coexistence. The extent of the *P. fragi*/*P. vulgaris* clade was delineated and sequence features specifically conserved in each of the *P. fragi*/*P. vulgaris* clade and other subfamily I.1 lipases were identified. The *P. shioyasakiensis* clade was newly identified as a group of lipases, which do not coexist with Lif genes in the genome; further, we showed that these lipases could be actively expressed in *E. coli* in a Lif-independent manner. Notably, the Lif-dependence of any given lipase can be easily predicted by examining the presence or absence of the conserved motifs suggested to be associated with Lif-dependence in this study. Overall, the results of this study contribute to expansion of the diversity of Lif-independent lipases available for industrial use. This study also provides fundamental insights into the evolution of lipases and Lif.

Abbreviations

Lif	Lipase-specific foldase
ORF	Open reading frame
HMM	Hidden Markov model
PAL	*Pseudomonas aeruginosa* lipase
PML	*Proteus mirabilis* lipase
PBD	Protein data bank
IPTG	Isopropyl β-D-thiogalactopyranoside

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11717-4.

Supplementary Material 1.
Supplementary Material 2.
Supplementary Material 3.
Supplementary Material 4.
Supplementary Material 5.
Supplementary Material 6.

Acknowledgements

We would like to thank Editage (www.editage.jp) for English language editing.

Author's contributions

TH conceived the study, conducted the analyses and experiments, analyzed the results, and wrote the manuscript.

Funding

Not applicable.

Data availability

The accession numbers for the protein sequences analyzed in this study are listed in Fig. 4, S1, Table 1, S1-4. The sequence data are deposited in the NCBI database and can be accessed at https://www.ncbi.nlm.nih.gov/.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

TH is the inventor on patent number JP2024118710A related to the proteins described in this work.

Received: 22 March 2025 Accepted: 15 May 2025
Published online: 22 May 2025

References

1. Jaeger KE, Ransac S, Dijkstra BW, Colson C, van Heuvel M, Misset O. Bacterial lipases. FEMS Microbiol Rev. 1994;15:29–63.
2. Yao W, Liu K, Liu H, Jiang Y, Wang R, Wang W, Wang T. A valuable product of microbial cell factories: microbial lipase. Front Microbiol. 2021;12:743377.
3. Arpigny JL, Jaeger KE. Bacterial lipolytic enzymes: classification and properties. Biochem J. 1999;343(1):177–83.
4. Hitch TCA, Clavel T. A proposed update for the classification and description of bacterial lipolytic enzymes. PeerJ. 2019;7:e7249.
5. Glogauer A, Martini VP, Faoro H, Couto GH, Müller-Santos M, Monteiro RA, et al. Identification and characterization of a new true lipase isolated through metagenomic approach. Microb Cell Fact. 2011;10:54.
6. Rosenau F, Tommassen J, Jaeger KE. Lipase-specific foldases. Chembiochem. 2004;5:152–61.
7. Sandkvist M. Biology of type II secretion. Mol Microbiol. 2001;40:271–83.
8. El Khattabi M, Van Gelder P, Bitter W, Tommassen J. Role of the lipase-specific foldase of *Burkholderia glumae* as a steric chaperone. J Biol Chem. 2000;275:26885–91.
9. Shibata H, Kato H, Oda J. Calcium ion-dependent reactivation of a *Pseudomonas* lipase by its specific modulating protein. LipB J Biochem. 1998;123:136–41.
10. El Khattabi M, Ockhuijsen C, Bitter W, Jaeger KE, Tommassen J. Specificity of the lipase-specific foldases of gram-negative bacteria and the role of the membrane anchor. Mol Gen Genet. 1999;261:770–6.
11. Oshima-Hirayama N, Yoshikawa K, Nishioka T, Oda J. Lipase from *Pseudomonas aeruginosa*. Production in *Escherichia coli* and activation in vitro with a protein from the downstream gene. Eur J Biochem. 1993;215:239–46.
12. Quyen DT, Schmidt-Dannert C, Schmid RD. High-level formation of active *Pseudomonas cepacia* lipase after heterologous expression of the encoding gene and its modified chaperone in *Escherichia coli* and rapid in vitro refolding. Appl Environ Microbiol. 1999;65:787–94.
13. Quyen DT, Le Giang TT, Nguyen TT, Oh TK, Lee JK. High-level heterologous expression and properties of a novel lipase from *Ralstonia* sp. M1. Protein Expr Purif. 2005;39:97–106.
14. Ogino H, Katou Y, Akagi R, Mimitsuka T, Hiroshima S, Gemba Y, et al. Cloning and expression of gene, and activation of an organic solvent-stable lipase from *Pseudomonas aeruginosa* LST-03. Extremophiles. 2007;11:809–17.
15. Wu X, You P, Su E, Xu J, Gao B, Wei D. In vivo functional expression of a screened *P*. aeruginosa chaperone-dependent lipase in *E. coli*. BMC Biotechnol. 2012;12:58.
16. Quyen TD, Vu CH, Le GTT. Enhancing functional production of a chaperone-dependent lipase in *Escherichia coli* using the dual expression cassette plasmid. Microb Cell Fact. 2012;11:29.
17. Pulido IY, Prieto E, Pieffet GP, Mendez L, Jimenez-Junca CA. Functional heterologous expression of mature lipase LipA from *Pseudomonas aeruginosa* PSA01 in *Escherichia coli* SHuffle and BL21 (DE3): effect of the expression host on thermal stability and solvent tolerance of the enzyme produced. Int J Mol Sci. 2020;21:3925.
18. Aoyama S, Yoshida N, Inouye S. Cloning, sequencing and expression of the lipase gene from Pseudomonas fragi IFO-12049 in *E. coli*. FEBS Lett. 1988;242:36–40.
19. Kim HK, Lee JK, Kim H, Oh TK. Characterization of an alkaline lipase from *Proteus vulgaris* K80 and the DNA sequence of the encoding gene. FEMS Microbiol Lett. 1996;135:117–21.
20. Lee H-W, Yoon S-J, Kim H-K, Park K-M, Oh T-K, Jung J-K. Overexpression of an alkaline lipase gene from *Proteus vulgaris* K80 in *Escherichia coli* BL21/pKLE. Biotechnol Lett. 2000;22:1543–7.
21. Alquati C, De Gioia L, Santarossa G, Alberghina L, Fantucci P, Lotti M. The cold-active lipase of *Pseudomonas fragi*. Heterologous expression, biochemical characterization and molecular modeling. Eur J Biochem. 2002;269:3321–8.
22. Sullivan ER, Leahy JG, Colwell RR. Cloning and sequence analysis of the lipase and lipase chaperone-encoding genes from *Acinetobacter calcoaceticus* RAG-1, and redefinition of a proteobacterial lipase family and an analogous lipase chaperone family. Gene. 1999;230:277–86.
23. Madalozzo AD, Martini VP, Kuniyoshi KK, de Souza EM, Pedrosa FO, Glogauer A, et al. Immobilization of LipC12, a new lipase obtained by metagenomics, and its application in the synthesis of biodiesel esters. J Mol Cat B Enzym. 2015;116:45–51.
24. Korman TP, Sahachartsiri B, Charbonneau DM, Huang GL, Beauregard M, Bowie JU. Dieselzymes: development of a stable and methanol tolerant lipase for biodiesel production by directed evolution. Biotechnol Biofuels. 2013;6:70.
25. Korman TP, Bowie JU. Crystal structure of *Proteus mirabilis* lipase, a novel lipase from the *Proteus*/psychrophilic subfamily of lipase family I. PLOS One. 2012;7(1):e52890.
26. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
27. Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2021;49:D10-7.
28. Schuler GD, Epstein JA, Ohkawa H, Kans JA. Entrez: molecular biology database and retrieval system. Methods Enzymol. 1996;266:141–62.
29. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150–2.
30. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. MAFFT-DASH: integrated protein sequence and structural alignment. Nucleic Acids Res. 2019;47:W5-10.
31. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4.
32. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49:W293-6.
33. Teufel F, AlmagroArmenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022;40:1023–5.
34. Wohlfarth S, Hoesche C, Strunk C, Winkler UK. Molecular genetics of the extracellular lipase of *Pseudomonas aeruginosa* PAO1. J Gen Microbiol. 1992;138:1325–35.
35. Ogierman MA, Fallarino A, Riess T, Williams SG, Attridge SR, Manning PA. Characterization of the *Vibrio cholerae* El Tor lipase operon *lipAB* and a protease gene downstream of the *hly* region. J Bacteriol. 1997;179:7072–80.
36. Kok RG, van Thor JJ, Nugteren-Roodzant IM, Vosman B, Hellingwerf KJ. Characterization of lipase-deficient mutants of *Acinetobacter calcoaceticus* BD413: identification of a periplasmic lipase chaperone essential for the production of extracellular lipase. J Bacteriol. 1995;177:3295–307.
37. Novototskaya-Vlasova K, Petrovskaya L, Kryukova E, Rivkina E, Dolgikh D, Kirpichnikov M. Expression and chaperone-assisted refolding of a new cold-active lipase from *Psychrobacter cryohalolentis* K5(T). Protein Expr Purif. 2013;91:96–103.
38. An SY, Kim SW, Park YL, Joo WH, Lee YC. Cloning, expression in *Escherichia coli*, and enzymatic properties of a lipase from *Pseudomonas* sp. SW-3. J Microbiol. 2003;41:95–101.
39. Dieckelmann M, Johnson LA, Beacham IR. The diversity of lipases from psychrotrophic strains of *Pseudomonas*: a novel lipase from a highly lipolytic strain of *Pseudomonas fluorescens*. J Appl Microbiol. 1998;85:527–36.
40. Ji X, Li S, Wang B, Zhang Q, Lin L, Dong Z, et al. Expression, purification and characterization of a functional, recombinant, cold-active lipase (LipA) from psychrotrophic *Yersinia enterocolitica*. Protein Expr Purif. 2015;115:125–31.
41. Frenken LG, Bos JW, Visser C, Müller W, Tommassen J, Verrips CT. An accessory gene, *lipB*, required for the production of active *Pseudomonas glumae* lipase. Mol Microbiol. 1993;9:579–89.
42. Quyen DT, Nguyen TT, Le TT, Kim HK, Oh TK, Lee JK. A novel lipase/chaperone pair from *Ralstonia* sp. M1: analysis of the folding interaction and evidence for gene loss in *R. solanacearum*. Mol Genet Genomics. 2004;272:538–49.
43. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44:D279-85.
44. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011;39:W29-37.
45. Blin K. ncbi-genome-download. Scripts to download genomes from NCBI FTP servers. https://github.com/kblin/ncbi-genome-download. Accessed 20 Mar 2022.
46. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14:1188–90.

47. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer. Preprint at bioRxiv. 2021. https://doi.org/10.1101/2021.10.04.463034.
48. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nat Methods. 2022;19:679–82.
49. Martínez A, Ostrovsky P, Nunn DN. LipC, a second lipase of *Pseudomonas aeruginosa*, is LipB and Xcp dependent and is transcriptionally regulated by pilus biogenesis components. Mol Microbiol. 1999;34:317–26.
50. Liebeton K, Zacharias A, Jaeger KE. Disulfide bond in *Pseudomonas aeruginosa* lipase stabilizes the structure but is not required for interaction with its foldase. J Bacteriol. 2001;183:597–603.
51. Nardini M, Lang DA, Liebeton K, Jaeger KE, Dijkstra BW. Crystal structure of pseudomonas aeruginosa lipase in the open conformation: the prototype for family I.1 of bacterial lipases. J Biol Chem. 2000;275:31219–25.
52. Pauwels K, Lustig A, Wyns L, Tommassen J, Savvides SN, Van Gelder P. Structure of a membrane-based steric chaperone in complex with its lipase substrate. Nat Struct Mol Biol. 2006;13:374–5.
53. de Oliveira Silva YR, Contreras-Martel C, Rodrigues de Melo R, Zanphorlin LM, Trindade DM, Dessen A. Architecture of an embracing lipase-foldase complex of the type II secretion system of *Acinetobacter baumannii*. Structure. 2025;33:601–12.
54. Kim EK, Jang WH, Ko JH, Kang JS, Noh MJ, Yoo OJ. Lipase and its modulator from Pseudomonas sp. strain KFCC 10818: proline-to-glutamine substitution at position 112 induces formation of enzymatically active lipase in the absence of the modulator. J Bacteriol. 2001;183:5937–41.

Publisher's Note